

# Oblique Random Forests for 3-D Vessel Detection Using Steerable Filters and Orthogonal Subspace Filtering\*

Matthias Schneider<sup>1</sup>, Sven Hirsch<sup>1</sup>, Gábor Székely<sup>1</sup>, Bruno Weber<sup>2</sup>,  
and Bjoern H. Menze<sup>1</sup>

<sup>1</sup> Computer Vision Laboratory, ETH Zurich, Switzerland

<sup>2</sup> Institute of Pharmacology and Toxicology, University of Zurich, Zurich, Switzerland

**Abstract.** We propose a machine learning-based framework using oblique random forests for 3-D vessel segmentation. Two different kinds of features are compared. One is based on orthogonal subspace filtering where we learn 3-D eigenspace filters from local image patches that return task optimal feature responses. The other uses a specific set of steerable filters that show, qualitatively, similarities to the learned eigenspace filters, but also allow for explicit parametrization of scale and orientation that we formally generalize to the 3-D spatial context. In this way, steerable filters allow to efficiently compute oriented features along arbitrary directions in 3-D. The segmentation performance is evaluated on four 3-D imaging datasets of the murine visual cortex at a spatial resolution of 0.7  $\mu\text{m}$ . Our experiments show that the learning-based approach is able to significantly improve the segmentation compared to conventional Hessian-based methods. Features computed based on steerable filters prove to be superior to eigenfilter-based features for the considered datasets. We further demonstrate that random forests using oblique split directions outperform decision tree ensembles with univariate orthogonal splits.

**Keywords:** vessel segmentation, orthogonal subspace filtering, steerable filters, oblique random forest.

## 1 Introduction

Blood vessel enhancement and segmentation play a crucial role for numerous medically oriented applications and has attracted a lot of attention in the field of medical image processing. The multiscale nature of vessels, image noise and contrast inhomogeneities make it a challenging task. In this context, a large variety of methods have been developed exploiting photometric and structural properties of tubular structures. Extensive reviews on various state-of-the-art vessel segmentation techniques can be found in the literature [14,15]. Rather simple methods, e.g., absolute or locally adaptive thresholding, are in fact regularly used in practice due to their conceptual simplicity and computational efficiency but they are a serious source of error and require careful parameter selection [20,22]. More sophisticated segmentation techniques such as optimal filtering and Hessian-based approaches commonly rely on idealized appearance

---

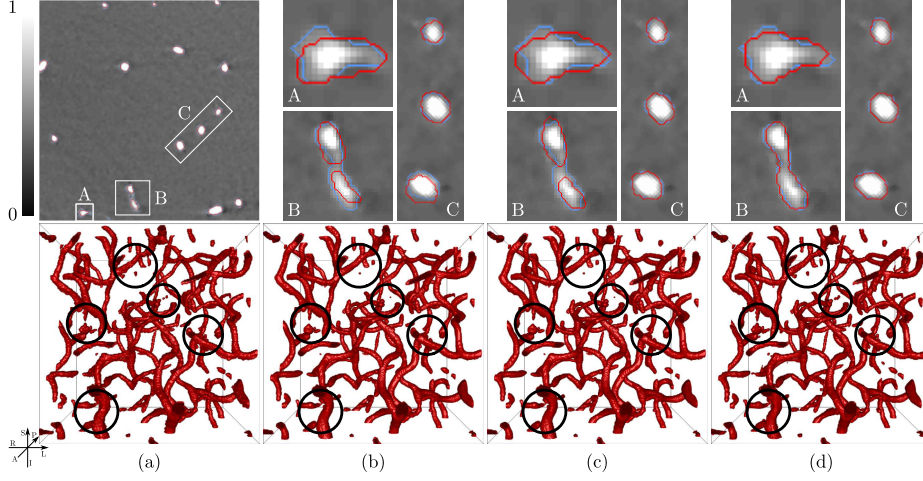
\* Supplementary material for this article is available at  
<http://www.vision.ee.ethz.ch/ReCoVa>

and noise models. The former includes optimal edge detection [2], and steerable filters providing an elegant theory for computationally efficient ridge detection at arbitrary orientations [12,9]. The latter is based on the eigenanalysis of the Hessian capturing the second order structure of local intensity variations [4,24]. The Hessian is commonly computed by convolving the image patch with the partial second order derivatives of a Gaussian kernel as the method of choice for noise reduction and to tune the filter response to a specific vessel scale. This basic principle has already been used by Canny for edge and line detection [2]. The differential operators involved in the computation of the Hessian are well-posed concepts of linear scale-space theory [16]. Modeling vessels as elongated elliptical structures, the eigendecomposition of the Hessian has a geometric interpretation, which can be used to define a “vesselness” measure as a function of the eigenvalues [4,24]. Due to the multi-scale nature of vascular structures, Hessian-based filters are commonly applied at different scales. Besides, the eigenvector corresponding to the largest eigenvalue of the Hessian computed at the most discriminative scale is a good estimate for the local vessel direction. In practice, vesselness filters tend to be prone to noise and have difficulty in detecting vessel parts such as bifurcations not complying with the intrinsic idealized appearance model. Vesselness filters have also been successfully applied for global vessel segmentation in X-ray angiography using ridge tracking [26] and graph cut theory [10].

In this paper, we devise a machine learning approach for vessel segmentation based on the 2-D filament detection framework proposed by Gonzalez et al. [9] using steerable filters [5,12]. In our application, we aim at efficient classification of 3-D high-resolution imaging datasets ( $> 10^{10}$  voxels) of the murine visual cortex (see Figure 1), which is of great interest for the analysis of the cerebrovascular system [22,11]. Due to the considerable computational challenge that comes with our application, we focus on a fast classification approach using local linear filters rather than complex non-local spatial models incorporating prior knowledge and regularization [26,10]. We compare different features computed from, respectively, orthogonal subspace filtering [17,23] and steerable filters using Gaussian derivatives [5,8]. In contrast to the framework proposed by Gonzalez et al. [9,8], we use oblique random forests (RF) for efficient classification. We test “elastic net” node models that combine  $\ell_1$  and  $\ell_2$  regularization leading to sparser node models than the  $\ell_2$  regularized oblique splits proposed in [18].

## 2 Methods

In this section, we first introduce two different sets of features based on (1) orthogonal subspace filtering and (2) steerable filters computed at different scales and orientations in order to achieve rotational invariance. These features are then used to train an oblique random forest (RF) classifier that is well adapted to correlated feature responses from local image filters [18]. Different from standard discriminative learning algorithms such as support vector machines, RF classifiers return continuous probabilities when predicting vessel locations, which allows to choose an operating point by adapting the decision threshold. Moreover, RF is capable of coping with high dimensional feature vectors and tolerate false training labels. It is fast to train with only very few parameters to be optimized and even faster to apply. Efficient prediction becomes particularly important in view of our specific application using high-resolution image data at  $\mu\text{m}$  resolution.



**Fig. 1.** Visualization of segmented cerebrovascular network for single axial slice (top) and whole 3-D test ROI (bottom) using different segmentation techniques. (a) Ground truth. (b) Frangi [4]. (c) RF-OSF ( $d = 102$ ). (d) RF-SFT ( $M = 4$ ). The binary segmentation maps are computed at the corresponding  $F_1$ -optimal operating points marked in Figure 4(b). The results are rendered in 3-D (bottom) and outlined in red (top) along with the ground-truth contours in blue for three subregions within the axial slice (A-C). Red contours in (a) mark the Otsu labels [20] used for RF training. Black circles in the 3-D plots highlight prominent differences in the segmentation. More results for the other datasets are provided in the supplementary material.

## 2.1 Orthogonal Subspace Filters (OSF)

Matched filters (MF) have widely been used in signal processing. They allow to detect a signal of known shape (template) by cross-correlation and perform provably optimal under additive Gaussian white noise conditions [19]. In terms of image processing, this corresponds to the convolution of the image with the MF. From a learning and classification perspective, matched filtering (signal detection) is closely related to linear regression for binary classification between background and pattern (vessel) [17]. Considering the image as a composition of local image patches with each pixel in the patch representing a feature, MF defines a 1-D linear subspace (regression coefficients) of this feature space which allows for separation of the pattern from background. Instead of an optimal 1-D subspace assuming linear separability in the feature space as implied by using a single matched filter, we use a less restrictive dimensionality reduction similar to [17], namely (linear) principal component analysis (PCA), in order to define a subspace of higher dimensionality. More formally, let  $\mathbf{p}_i \in \mathbb{R}^{P^3}$  denote a (cubic) image patch of size  $P \times P \times P$ . A  $d$ -dimensional subspace ( $d \leq P^3$ ) capturing the most important modes of variation in the image patches can then be defined using PCA [13]:

$$\forall 1 \leq k \leq d \leq P^3 : \boldsymbol{\alpha}_k = \arg \max_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{P^3}, \|\boldsymbol{\alpha}\|=1, \\ \forall 1 \leq i < k: \text{cov}(\boldsymbol{\alpha}_i, \boldsymbol{\alpha})=0}} \text{var}(\boldsymbol{\alpha}^T P_{\text{OSF}}) \quad , \quad (1)$$

where  $P_{\text{OSF}} = [\mathbf{p}_i]_{1 \leq i \leq N_P} \in \mathbb{R}^{P^3 \times N_P}$  is the data matrix assembling  $N_P$  patches labeled as vessel. The principal axes  $\boldsymbol{\alpha}_k$  form an orthonormal basis of the  $d$ -dimensional subspace and are ordered according to their preserved variance. They can be computed efficiently as the  $d$  eigenvectors corresponding to the largest eigenvalues of the covariance matrix of  $P_{\text{OSF}}$  after mean centering using singular value decomposition. Projecting an arbitrary image patch  $\mathbf{p} \in \mathbb{R}^{P^3}$  onto the PCA subspace yields its  $d$  principal components (PC). The PCs of the image patches centered at pixels  $\mathbf{x}$  in image  $I$  can thus be computed by  $d$  independent convolution operations of the image with each (properly reshaped) principal axis  $\tilde{\boldsymbol{\alpha}}_k \in \mathbb{R}^{P \times P \times P}$ :

$$\mathbf{f}_{\text{OSF}}(I, \mathbf{x}) = \left[ (\tilde{\boldsymbol{\alpha}}_k * I)(\mathbf{x}) - \boldsymbol{\alpha}_k^T \frac{1}{N_P} \sum_{i=1}^{N_P} \mathbf{p}_i \right]_{1 \leq k \leq d} \in \mathbb{R}^d \quad . \quad (2)$$

The (reshaped) principal axes will also be referred to as orthogonal subspace filters (OSF). The PCs, i.e., the OSF response of an image patch, are used as features along with a non-linear decision rule for vessel segmentation as described in Section 2.3

## 2.2 Steerable Filter Templates (SFT)

The OSF eigenfilters learned from image patches as described in the previous section turn out to be highly structured (see Figure 2(a)). Instead of learning the structured filter kernels, we hence attempt to explicitly parametrize them. For this, we choose a steerable filter model based on Gaussian derivatives, which allows for efficient directional filtering at different scales and, most importantly, implicates rotational invariance [12]. Similar to [8], we define the filter templates as normalized derivatives of Gaussians up to order  $M$  [16]:

$$\forall m \geq 1 \wedge 0 \leq b \leq a \leq m \leq M : G_{m,a,b}^\sigma(\mathbf{x}) = \sigma^m \frac{\partial^{m-a} \partial^{a-b} \partial^b}{\partial x^{m-a} \partial y^{a-b} \partial z^b} G^\sigma(\mathbf{x}) \quad , \quad (3)$$

where  $G^\sigma(\mathbf{x}) = \frac{1}{(\sqrt{2\pi}\sigma)^3} \exp(-\frac{\|\mathbf{x}\|^2}{2\sigma^2})$  denotes the 3-D symmetric Gaussian kernel with variance  $\sigma$  and zero mean. As in Equation (2), each template induces a single feature by convolution with image  $I$ . They can be assembled to a feature vector of dimension  $d_M = 1/6(M^3 + 6M^2 + 11M)$  at a fixed scale  $\sigma$ :

$$\mathbf{f}^\sigma(I, \mathbf{x}) = ((G_{1,0,0}^\sigma, G_{1,1,0}^\sigma, G_{1,1,1}^\sigma, \dots, G_{M,M,M}^\sigma)^T * I)(\mathbf{x}) \in \mathbb{R}^{d_M} \quad . \quad (4)$$

We enhance the features by concatenating feature vectors at different scales  $\sigma_1, \dots, \sigma_S$ :

$$\mathbf{f}_{\text{SFT}}(I, \mathbf{x}) = (\mathbf{f}^{\sigma_1}(I, \mathbf{x}), \dots, \mathbf{f}^{\sigma_S}(I, \mathbf{x}))^T \in \mathbb{R}^{d_M S} \quad . \quad (5)$$

The steerability of Gaussian derivatives has been derived for the 2-D case in [12] and can readily be extended to 3-D [58]. Steerability refers to the property that the convolution of an image with a rotated version of the steerable filter template (SFT) can

be expressed by a linear combination of the filter response of the image with the SFT without rotation:

$$I * G_{m,a,b}^\sigma(R\mathbf{x}) = \sum_{i=0}^m \sum_{j=0}^i \omega_{m,a,b}^{i,j} \underbrace{(I * G_{m,i,j}^\sigma)(\mathbf{x})}_{\mathbf{f}_{m,i,j}^\sigma(I,\mathbf{x})} \quad , \quad (6)$$

where  $R \in SO(3)$  denotes a 3-D rotation matrix and  $\omega_{m,a,b}^{i,j}$  the uniquely defined coefficients that can be computed in closed form [12]<sup>1</sup>. This formalism allows to efficiently evaluate the feature vector  $\mathbf{f}_{\text{SFT}}$  for an arbitrary rotation without any additional costly convolution. We use a restricted set of rotations in our application considering the tubular structure of vessels. The local vessel direction  $\mathbf{d} = (d_x, d_y, d_z)^T \in \mathbb{R}^3$ ,  $\|\mathbf{d}\| = 1$  can be parametrized using spherical coordinates  $(\theta, \phi)$  with unit radius, elevation  $\theta = \arctan(d_z / \sqrt{d_x^2 + d_y^2})$ , and azimuth  $\phi = \arctan(d_y / d_x)$  relative to the  $x$ - $y$  plane ( $z = 0$ ). It is sufficient to restrict the parametrization to the positive hemisphere ( $z > 0$ ), i.e.,  $0 \leq \theta \leq \pi/2$  and  $-\pi < \phi \leq \pi$ . The vessel can then be transformed to the normalized pose  $\mathbf{d}_0 = (1, 0, 0)^T$  by applying the rotation matrix

$$R_{\theta,\phi} = \begin{pmatrix} \cos \theta \cos \phi & \cos \theta \sin \phi & \sin \theta \\ -\sin \phi & \cos \phi & 0 \\ -\sin \theta \cos \phi & -\sin \theta \sin \phi & \cos \theta \end{pmatrix} \quad . \quad (7)$$

The SFT features evaluated for this rotation according to Equation (6) hence describe the intensity variation characteristics of different order along the vascular structure as well as in the orthogonal plane. Assuming a symmetric vessel (intensity) profile perpendicular to the local vessel direction  $\mathbf{d}$ , restricting the set of rotations is reasonable as the vessel appearance is (locally) invariant under rotation about  $\mathbf{d}$ .

### 2.3 Vessel Classification - Shape Learning and Prediction

The OSF and SFT features as defined in Equations (2) and (5), respectively, are each used along with a non-linear decision rule for vessel segmentation. We train separate classifiers for the different feature types as follows: A representative set  $\mathcal{S}$  of  $2N_S$  tuples (image  $I_k$ , location  $\mathbf{x}_k$ , vessel orientation  $\mathbf{d}_k$ , class label  $y_k$ ) is randomly sampled from a labeled set of images corresponding to  $N_S$  foreground ( $y_k = 1$ ) and background ( $y_k = -1$ ) samples, respectively:  $\mathcal{S} = \{(I_k, \mathbf{x}_k, \mathbf{d}_k, y_k) \mid 1 \leq k \leq 2N_S\}$ . For these samples, the features  $\mathbf{f}(I, \mathbf{x})$  can be extracted as defined in Equations (2) and (5). The SFT features are additionally rotated to the normalized orientation according to Equations (6) and (7) w.r.t. the local vessel direction  $\mathbf{d}$ . This defines the training set  $\mathcal{T} = \{(\mathbf{f}_k = \mathbf{f}(I_k, \mathbf{x}_k), y_k) \mid 1 \leq k \leq 2N_S\}$  that is ultimately used to train a random forest (RF) classifier [1]. RF consists of an ensemble of decision trees used to model the posterior probability of each class (vessel/background). During training, each tree is fully grown from bootstrapped datasets using stochastic discrimination. For this, the data is split at each tree node by a hyperplane in the feature (sub-)space. In contrast to traditional bagging, the split is based on a small number of randomly selected

<sup>1</sup> Further details are provided in the supplementary material.

features only. We investigated both “orthogonal” and “oblique” trees. As proposed in Breiman’s original paper [1], the former is based on optimal thresholds for randomly selected single features in every split, i.e., mutually orthogonal 1-D hyperplanes. The latter uses multidimensional hyperplanes to separate the feature space, e.g., by choosing randomly oriented hyperplanes [1] or applying linear discriminative models [18]. For the oblique RFs in this work, we employ a linear regression with an elastic net penalty [6] in order to learn multivariate (optimal) split directions  $\mathbf{w}$  at each node:

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w} \in \mathbb{R}^{N_F}} \frac{1}{2|\mathcal{T}|} \sum_{k=1}^{|\mathcal{T}|} \left( y_k - \mathbf{w}^T \tilde{\mathbf{f}}_k \right)^2 + \lambda P_\alpha(\mathbf{w}) \quad , \quad (8)$$

where  $\tilde{\mathbf{f}}_k \in \mathbb{R}^{N_F}$  are randomly selected (but fixed) features and  $\lambda > 0$  is the regularization parameter for the elastic net penalty  $P_\alpha(\mathbf{w}) = (1 - \alpha)\frac{1}{2}\|\mathbf{w}\|_{\ell_2}^2 + \alpha\|\mathbf{w}\|_{\ell_1}$  as a compromise between the ridge regression ( $\alpha = 0$ ) and the lasso penalty ( $\alpha = 1$ ), where  $\|\cdot\|_{\ell_1}$  and  $\|\cdot\|_{\ell_2}$  denote the  $\ell_1$  and  $\ell_2$ -norm, respectively. The advantage is joint regularization of the coefficients and sparsity — coefficients are both encouraged to be small, and to be zero if they are very small. The latter lasso property reduces the dimensionality of the split space, which is desirable for memory and robustness purposes. With  $\alpha = 1$  (and  $\lambda \gg 0$ ) we will get a single non-zero coefficient, i.e., RF with univariate splits, whereas choosing  $\alpha = 0$  we have ridge regression as in [18].

The decision trees are grown separately as follows:

1. For each tree, a new set of samples is randomly drawn from the training data  $\mathcal{T}$  with replacement, i.e.,  $\frac{2}{3}|\mathcal{T}|$  bootstrapped samples.
2. For every node,  $N_F$  features are randomly sampled without replacement from the feature pool of size  $N_F^0 = d$  for OSF features and  $N_F^0 = d_M S$  for SFT features, respectively (see Equations (2) and (4)).
3. The selected features of the bootstrapped samples are normalized to zero mean and unit variance at every split in order to enhance the stability of the linear model.
4. Finding optimal split
  - a) Orthogonal split ( $N_F = 1$ ): The feature values of all samples are tested as threshold to split the data w.r.t. the selected feature.
  - b) Oblique split ( $N_F = \lceil \sqrt{N_F^0} \rceil$ ): The optimal split direction is computed according to Equation (8) for  $\alpha = 0.5$  using covariance updates [6].
5. Steps 2–4 are repeated  $\lceil \sqrt{N_F^0} \rceil$  times. The optimal split and threshold are ultimately selected w.r.t. the information gain as a result of the split. The samples are split accordingly and passed on to the child nodes.
6. For each of the  $N_T$  trees, steps 2–5 are repeated until (1) all samples in a (leaf) node belong to the same class, (2) the maximum tree depth has been reached, or (3) there are too few samples to further split the data (avoid excessive overfitting).
7. Each leaf node is assigned a class label according to the majority vote of the training samples ending up in the considered leaf.

Previously unseen samples (images) can be classified by pushing the extracted features down all  $N_T$  decision trees of the ensemble. Thus, each tree assigns a class label

$\hat{y}_i \in \{0, 1\}$  associated with the leaf node in which the tested sample ends up. The ensemble confidence can then be defined as  $\frac{1}{N_T} \sum_{i=1}^{N_T} \hat{y}_i$  as an estimate of the posterior. The binary class label  $\hat{y}$  can finally be assigned using a majority vote or any other decision threshold.

In the case of OSF features, a single RF is trained for all vessel orientations. Therefore, the intrinsic orientation-induced structure in the OSF feature space has to be captured in the training set both for RF training and learning the OSF eigenfilters. In contrast, SFT features allow for explicit parametrization of the orientation. The expected filter response for an arbitrary orientation can efficiently be computed from the set of stationary base features  $\mathbf{f}_{\text{SFT}}$  as defined in Equations (5) and (6). As the corresponding RF classifiers are trained on SFT features extracted from vessels with normalized orientation only, we sample the space of possible vessel orientations (half sphere) and compute the corresponding (rotated) SFT features in order to build an orientation independent predictor. The classification result with the maximum confidence is ultimately assigned as proposed in [9]. In contrast to OSF features, this allows to not only estimate the class posteriors but also a probability distribution on the vessel orientation.

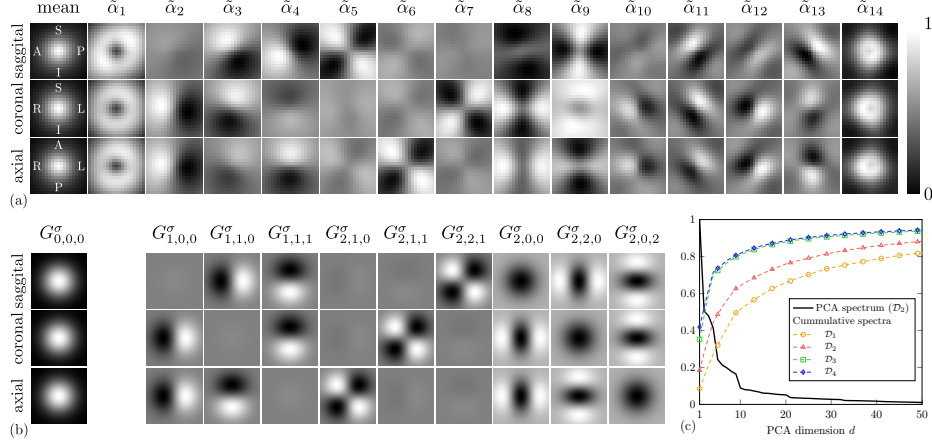
### 3 Experiments

We have evaluated the performance of our method on four 3-D datasets  $\mathcal{D}_{1-4}$  obtained from synchrotron radiation X-ray tomographic microscopy (srXTM) of cylindrical samples of the murine somatosensory cortex (volume size  $2048 \text{ px} \times 2048 \text{ px} \times 4000 \text{ px}$ , isotropic voxel spacing  $0.7 \mu\text{m}$ , grayscale 16 bit) [22]. In a preprocessing step we applied anisotropic diffusion filtering in order to reduce image noise while preserving edge contrast [21]. From each (preprocessed) dataset we extracted two disjoint regions of interest (ROI) of size  $(256 \text{ px})^3$  for training and testing, respectively. In the following, we will refer to these non-overlapping ROIs as test and train data/ROI, respectively (see Figure 1(a)). For each test ROI, ground truth labels were manually generated by an expert assisted by a semi-automatic segmentation tool [27] on 15 evenly distributed slices along each reference direction (axial, coronal, sagittal). Thus, 125 slices have been labeled containing  $7.3 \times 10^4$  foreground and  $2.7 \times 10^6$  background labels in average ( $\pm 3.9 \times 10^4$ ) corresponding to a vascular volume fraction of  $2.6 \pm 1.4 \%$ .

In a first baseline experiment, all ROIs were segmented using both Otsu’s method [20] and multiscale vessel enhancement filtering [4, 24]. For the latter, we have performed an exhaustive grid search to optimize the vesselness scale on the test ROIs w.r.t. maximum area under the ROC curve using the ground-truth labels of the test ROIs. In the majority of the cases five logarithmically spaced scales performed best for both Frangi’s and Sato’s vesselness:  $\sigma \in \{2.00, 3.09, 4.76, 7.35, 11.33\} [\text{px}]$ .

In a next step, we computed the OSF eigenfilters introduced in Section 2.1 from 3000 randomly sampled patches centered at voxels labeled as vessel in the Otsu label map. In particular, background patches were not considered during OSF learning. Besides the original vessel patches, five randomly rotated versions of each patch have been added to the set of patches  $P_{\text{OSF}}$  used in Equation (1) in order to account for rotational symmetry of vessel structures while keeping the total number of patches at a moderate level ( $N_P = 1.8 \times 10^4$ ). As in [17], the OSF patch size  $P$  was assessed from the random forest feature importance and set to  $P = 19$ .





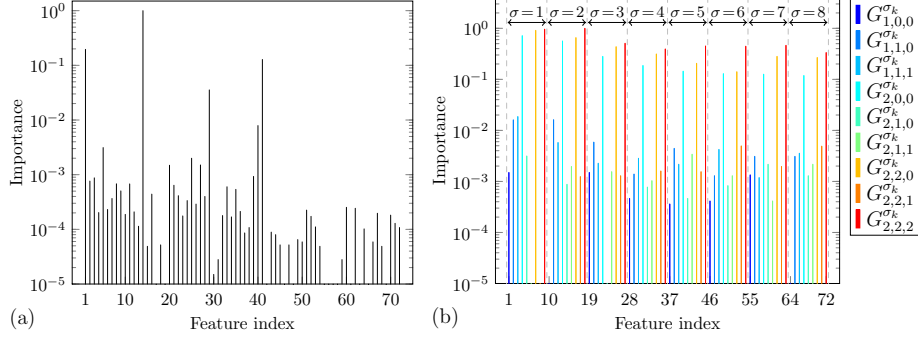
**Fig. 2.** (a) Visualization of the mean pattern and the most significant (reshaped) eigenfilters  $\tilde{\alpha}_k$  along centered sagittal, coronal, and axial slices as learned from dataset  $\mathcal{D}_2$  ( $P = 19$ ). (b) Normalized Gaussian derivatives  $G_{m,a,b}^\sigma$  at a fixed scale  $\sigma$  up to order  $M = 2$  as defined in Equation (3). (c) Normalized PCA spectrum  $\lambda_k/\lambda_1$  and variance preservation as measured by the cumulative spectrum  $\sum_{k=1}^d \lambda_k / \sum_{k=1}^{P^3} \lambda_k$  for different datasets, where  $\lambda_k$  denotes the  $k$ -th eigenvalue of the data covariance matrix.

As for the SFT model, we performed a small parameter study to optimize the SFT scales similar to the multiscale vesselness parameters. In order to avoid overfitting, however, we used the train ROIs for the parameter optimization along with the Otsu labels considered as ground truth in this case. We ultimately select  $S = 3$  logarithmically spaced scales  $\sigma \in \{2.00, 3.65, 6.67\}$ . The SFT model hence defines  $d_M S = 9$  (27, 57, 102) features for maximum Gaussian derivative order  $M = 1$  (2, 3, 4), respectively (see Equations (4) and (5)). For a fair comparison of the SFT and OSF feature models, the PCA subspace dimension  $d$  of the OSF models, i.e., the number of OSF features, was chosen accordingly.

Different RF classifiers consisting of  $N_T = 256$  decision trees have been trained separately on the train ROI of a single dataset using OSF and SFT features along with orthogonal and oblique splits, respectively, as explained in Section 2.3. The training was repeated for each dataset using  $N_S = 4000$  foreground (vessel) and background samples, respectively, randomly drawn from the Otsu label map. The local vessel direction was estimated from the eigenanalysis of the Hessian computed at the most discriminative scale as defined by Frangi's multiscale vesselness [4]. Note that the training labels were computed fully automatically without any user interaction. The manually annotated ground-truth labels have been used for RF validation only.

Finally, the different RF models were applied to the test ROIs of each dataset. The classification performance was evaluated on the uniformly aligned slices with ground-truth labels available (see above). In this way, the generalization of the individual





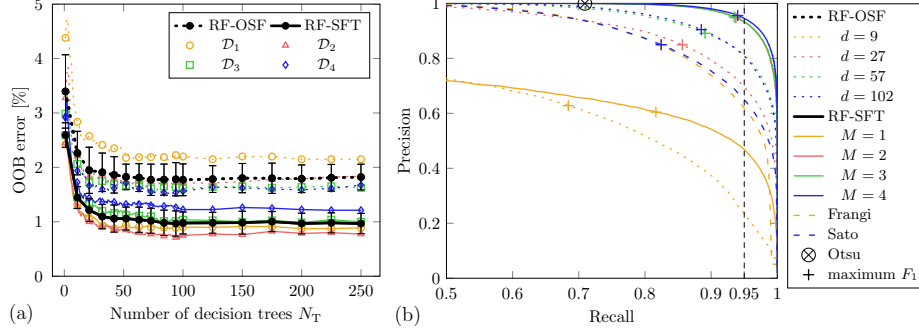
**Fig. 3.** Variable importance [11] of the (a) RF-OSF model ( $P = 19$ ,  $d = 57$ ) and (b) RF-SFT model ( $M = 2$ ,  $d_M = 9$ ,  $S = 8$  scales  $\sigma \in \{1, \dots, 8\}$ ) on a logarithmic scale (oblique splits). The prominent peaks in (b) correspond to the Gaussian derivatives  $G_{2,0,0}^\sigma$ ,  $G_{2,2,0}^\sigma$ , and  $G_{2,2,2}^\sigma$ .

classifiers is investigated (test ROIs of datasets not used for training) as well as the prediction quality for unseen samples from the dataset used for training (train ROI) but from a different subvolume (test ROI).

## 4 Results and Discussion

The learned OSF filter templates are highly structured (see Figure 2(a)). The ball-shaped mean shows a Gaussian-like pattern. The most significant principal axis captures the average image intensity in the vicinity of the sample. Patches  $\alpha_2, \dots, \alpha_4$  capture first order derivatives along the right-left (R-L), superior-inferior (S-I), and anterior-posterior (A-P) direction, respectively. Similar first-order patterns at a smaller scale appear in  $\alpha_{10}, \dots, \alpha_{13}$ . Differently oriented second order derivatives are described by  $\alpha_5, \dots, \alpha_9$ . The corresponding PCA spectra show a sharp profile as indicated in Figure 2(c). These observations can be made for all OSF models regardless of the considered patch size. For comparison of the structural similarities, the parameterized Gaussian derivatives up to order  $M = 2$  as used for the SFT feature extraction are shown in Figure 2(b).

The normalized RF feature relevance score, i.e., the permutation importance from [11], for the RF-OSF and RF-SFT model using oblique splits are shown in Figure 3. The OSF patches describing the average image intensity in the local neighborhood ( $\alpha_1$ ,  $\alpha_{14}$ ) show high variable importance as compared to the patches capturing higher order derivatives  $\alpha_2, \dots, \alpha_{13}$ . It also becomes clear that the OSF feature importance (discrimination capability) is not correlated to the PCA spectrum (variance preservation). This makes it difficult to choose a proper cutoff for the PCA subspace dimension. The variable importance of the SFT features indicates that the second order derivatives parallel and orthogonal to the vessel direction ( $G_{2,0,0}^\sigma$ ,  $G_{2,2,0}^\sigma$ ,  $G_{2,2,2}^\sigma$ ) are most significant for the classification. Note that the Hessian-based segmentation approaches also rely on these features [4,24]. For larger scales  $\sigma$ , the importance values tend to decline.



**Fig. 4.** Comparison of the classification performance. (a) Out of bag (OOB) error of the RF-OSF ( $d = 102$ ) and RF-SFT ( $M = 4$ ,  $d_M = 34$ ) classifiers trained on dataset  $\mathcal{D}_k$  for varying number of trees  $N_T$  (oblique splits). The average error is plotted in black with error bars indicating the standard deviation. (b) Precision-recall curves (PRC) and optimal operating points w.r.t.  $F_1$  measure for RF-OSF and RF-SFT models ( $N_T = 256$ , oblique splits, trained on  $\mathcal{D}_2$ ) with varying parameters  $d$  and  $M$ , respectively, in comparison to (optimized) Frangi’s/Sato’s vesselness, and Otsu thresholding [20] evaluated on test ROI of  $\mathcal{D}_2$ .

Figure 4(a) visualizes the out of bag (OOB) error of the RF-OSF and RF-SFT classifiers for different number of decision trees  $N_T$ . In both cases the OOB error declines rapidly for increasing  $N_T$ . The SFT model consistently shows smaller error rates compared to the OSF features. Moreover, the RF-SFT classifier is more robust across different datasets as indicated by the smaller standard deviation. Also note that the absolute values of the OOB error estimates may be somewhat overoptimistic due to the spatial correlation between the training samples.

Comparing the overall classification performance of the proposed learning-based approaches with different model parameters to standard segmentation approaches reveals the superior performance of the SFT features as indicated by the precision-recall curves (PRC) in Figure 4(b). The RF-based segmentation outperforms Frangi’s/Sato’s vesselness filters even for a small number of features ( $d = 27$ ,  $M = 2$ ). Note that the reported results for the vesselness-based segmentation have to be considered as upper bound as the scale parameters have been optimized on the test data (overfitting). The analysis also shows that for  $M > 1$  the performance of the RF-SFT model hardly changes anymore, which is consistent with the observation of the second order derivatives being the most discriminative features (see Figure 3(d)).

A more detailed numerical analysis of the classification performance of the different approaches is summarized in Table 1 and confirms the superior performance of the RF-SFT model over the OSF features and the multiscale vesselness filters. Otsu’s method [20] tends to underestimate the global threshold and hence results in an inaccurate segmentation of the vessel boundaries as indicated by the increased balanced error rate [3]. In order to assess the robustness of the learning-based segmentation approaches, we apply “intra-dataset” and “inter-dataset” cross-validation, i.e., choosing the (non-overlapping) train and test ROIs from the same (intra) or different (inter) datasets, respectively. The average segmentation performance for “totally” unseen data

**Table 1.** Detailed evaluation of classification performance of different RF-OSF ( $d = 102$ ) and RF-SFT ( $M = 4$ ,  $d_M = 34$ ) classifiers ( $N_T = 256$ ) using orthogonal and oblique splits, respectively. The performance is evaluated using “intra-dataset” and “inter-dataset” cross-validation (see text). The operating point was selected at the 95 % recall level (see Figure 4(b)). The partial area under the precision-recall curve (AUC-PR) has been computed on the recall interval  $[0.5, 1]$ .

	Method	Validation	Precision [%]	Specificity [%]	Error Rate [%]	AUC-PR [ $\times 10^{-2}$ ]	OOB Error [%]	Tree Depth
orthogonal	RF-OSF	intra-data	$74.32 \pm 7.26$	$99.20 \pm 0.13$	$2.92 \pm 0.06$	$45.29 \pm 1.74$	$2.02 \pm 0.44$	$9.02 \pm 0.45$
		inter-data	$70.99 \pm 7.55$	$98.97 \pm 0.62$	$3.06 \pm 0.33$	$44.35 \pm 2.09$		
	RF-SFT	intra-data	$89.43 \pm 1.19$	$99.70 \pm 0.14$	$2.70 \pm 0.09$	$48.35 \pm 0.20$	$1.35 \pm 0.22$	$7.42 \pm 0.38$
		inter-data	$88.25 \pm 2.05$	$99.67 \pm 0.14$	$2.70 \pm 0.07$	$48.13 \pm 0.41$		
oblique	RF-OSF	intra-data	$78.96 \pm 5.81$	$99.37 \pm 0.16$	$2.84 \pm 0.07$	$46.38 \pm 1.19$	$1.82 \pm 0.25$	$6.26 \pm 0.21$
		inter-data	$78.35 \pm 4.39$	$99.33 \pm 0.26$	$2.87 \pm 0.15$	$45.95 \pm 1.15$		
	RF-SFT	intra-data	$93.53 \pm 1.47$	$99.83 \pm 0.07$	$2.62 \pm 0.05$	$48.96 \pm 0.22$	$0.96 \pm 0.19$	$5.85 \pm 0.24$
		inter-data	$92.80 \pm 1.94$	$99.82 \pm 0.06$	$2.62 \pm 0.03$	$48.84 \pm 0.31$		
	Sato	average	$62.15 \pm 2.71$	$98.46 \pm 0.75$	$3.27 \pm 0.37$	$42.34 \pm 0.64$	n/a	n/a
	Frangi	average	$59.61 \pm 2.09$	$98.26 \pm 0.90$	$3.37 \pm 0.45$	$41.66 \pm 0.53$	n/a	n/a
	Otsu	average	$99.96 \pm 0.03$	$100.00 \pm 0.00$	$14.59 \pm 1.38$	n/a	n/a	n/a

(inter-dataset) slightly decreases compared to the (still unseen) test data in the case of intra-dataset validation. The figures also reveal that oblique splits, as compared to orthogonal splits, yield both better classification performance and smaller (average) tree depth. The advantage of oblique over orthogonal splits may result from the highly correlated features [18]. Further experiments would be required to investigate the influence of the elastic net penalty of Equation (8) in more detail.

Figure 1 compares the binary segmentation of the cerebrovascular networks for the different approaches applied to the test data  $\mathcal{D}_2$  using the  $F_1$ -optimal operating points marked in Figure 4(b). Visually, the Frangi filter and partly also the RF-OSF model generate very smooth networks missing some of the details on the vessel surface. The ideal elliptical appearance model underlying the Hessian-based vesselness filters produces many false negatives at bifurcations, in particular, where the model assumptions do not hold. Here the classification approach is able to consider more complex geometries, that are in accordance with higher order filter responses in the training data. As already indicated by the precision-recall analysis, the axial views reveal that the Frangi segmentation varies significantly from the ground-truth labels in many cases, whereas the RF-OSF and especially the RF-SFT results are in much better agreement to the reference segmentation.

## 5 Conclusions and Future Work

We have compared two kinds of features for 3-D vessel segmentation using a machine learning approach. Starting from orthogonal subspace filtering, we learn an orthogonal basis from vessel patches to describe the local vessel appearance in a low-dimensional feature space. In a second step, we parametrize and approximate the highly structured base filters by Gaussian derivatives, which allows to efficiently decompose the image into a multiscale rotational basis using steerable filter theory [12, 8]. Both kinds of features are used to train random forest classifiers for vessel segmentation. The steerable filters in fact allow to train a single classifier on normalized (canonically oriented) vessel

samples as proposed in [9] for 2-D filament detection. Our experiments on 3-D high-resolution srXTM imaging data of the murine visual cortex demonstrate that the steerable filter features outperform the orthogonal subspace features. Moreover, the machine learning approach proves to be superior to Hessian-based segmentation approaches, especially for vessel structures, such as bifurcations, that cannot easily be modeled explicitly and violate the common cylindrical appearance assumption. The RF classifiers show excellent classification performance on the 3-D datasets even for imperfect and incomplete training data as obtained by Otsu's method in our experiments. The proposed segmentation framework hence allows to fully automatically learn RF models for 3-D vessel segmentation on new datasets.

The choice of the type of splits to be used in the decision tree ensembles of the RF classifier turned out to have a major impact on the classification performance. For our task, oblique splits using linear regression are clearly favorable over univariate orthogonal splits. Besides a more comprehensive study on the choice of the elastic net penalty, it would be interesting to investigate if more complex information such as vessel caliber or centerline can be learned and predicted in a general and computationally cheap fashion on different types of 3-D angiographic datasets by extending the framework using Hough forests [7]. These additional data on the vessel morphology and topology may allow to ultimately reconstruct physiologically consistent full-fledged cerebrovascular networks possibly in combination with proper methods to replace or extend missing or faulty regions by synthetic vasculatures [25] in order to overcome shortcomings of the reconstruction technique or limitations of the imaging modality.

**Acknowledgements.** This work has been funded by the Swiss National Center of Competence in Research on Computer Aided and Image Guided Medical Interventions (NCCR Co-Me) supported by the Swiss National Science Foundation.

## References

1. Breiman, L.: Random forests. *Mach. Learn.* 45, 5–32 (2001)
2. Canny, J.: Finding edges and lines in images. Tech. rep., Massachusetts Institute of Technology, Cambridge, MA, USA (1983)
3. Chen, Y.W., Lin, C.J.: Combining SVMs with various feature selection strategies. In: Guyon, I., Nikravesh, M., Gunn, S., Zadeh, L. (eds.) *Feature Extraction. STUDFUZZ*, vol. 207, pp. 315–324. Springer, Heidelberg (2006)
4. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale Vessel Enhancement Filtering. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) *MICCAI 1998. LNCS*, vol. 1496, pp. 130–137. Springer, Heidelberg (1998)
5. Freeman, W.T., Adelson, E.H.: The design and use of steerable filters. *IEEE Trans. Pattern Anal. Mach. Intell.* 13(9), 891–906 (1991)
6. Friedman, J.H., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33(1), 1–22 (2010)
7. Gall, J., Yao, A., Razavi, N., Van Gool, L., Lempitsky, V.: Hough forests for object detection, tracking, and action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 33(11), 2188–2202 (2011)
8. González, G., Aguet, F., Fleuret, F., Unser, M., Fua, P.: Steerable Features for Statistical 3D Dendrite Detection. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) *MICCAI 2009, Part II. LNCS*, vol. 5762, pp. 625–632. Springer, Heidelberg (2009)

9. González, G., Fleurety, F., Fua, P.: Learning rotational features for filament detection. In: CVPR 2009, pp. 1582–1589 (June 2009)
10. Hernández-Vela, A., Gatta, C., Escalera, S., Igual, L., Martín-Yuste, V., Radeva, P.: Accurate and Robust Fully-Automatic QCA: Method and Numerical Validation. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 496–503. Springer, Heidelberg (2011)
11. Hirsch, S., Reichold, J., Schneider, M., Székely, G., Weber, B.: Topology and hemodynamics of the cortical cerebrovascular system. *J. Cereb. Blood Flow Metab* (April 2012)
12. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny-like criteria. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(8), 1007–1019 (2004)
13. Jolliffe, I.T.: *Principal Component Analysis*, 2nd edn. Springer (2002)
14. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. *ACM Comput. Surv.* 36, 81–121 (2004)
15. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: A review of 3D vessel lumen segmentation techniques: models, features and extraction schemes. *Med. Image Anal.* 13(6), 819–845 (2009)
16. Lindeberg, T.: Edge detection and ridge detection with automatic scale selection. *Int.J. Comput. Vis.* 30, 465–470 (1996)
17. Menze, B.H., Kelm, B.M., Hamprecht, F.A.: From eigenspot to fisherspot - latent spaces in the nonlinear detection of spot patterns in a highly varying background. In: Decker, R., Lenz, H.J. (eds.) *Advances in Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization.*, vol. 33, pp. 255–262. Springer (2006)
18. Menze, B.H., Kelm, B.M., Splithoff, D.N., Koethe, U., Hamprecht, F.A.: On Oblique Random Forests. In: Gunopulos, D., Hofmann, T., Malerba, D., Vazirgiannis, M. (eds.) *ECML PKDD 2011, Part II. LNCS*, vol. 6912, pp. 453–469. Springer, Heidelberg (2011)
19. Moon, T., Stirling, W.: *Mathematical methods and algorithms for signal processing*. Prentice Hall (2000)
20. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE T. Syst. Man Cyb.* 9(1), 62–66 (1979)
21. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 629–639 (1990)
22. Reichold, J., Stampanoni, M., Keller, A.L., Buck, A., Jenny, P., Weber, B.: Vascular graph model to simulate the cerebral blood flow in realistic vascular networks. *J. Cereb. Blood Flow Metab.* 29(8), 1429–1443 (2009)
23. Rigamonti, R., Türetken, E., González Serrano, G., Fua, P., Lepetit, V.: Filter learning for linear structure segmentation. Tech. rep., Swiss Federal Institute of Technology, Lausanne (EPFL) (2011)
24. Sato, Y., Nakajima, S., Atsumi, H., Koller, T., Gerig, G., Yoshida, S., Kikinis, R.: 3D Multi-Scale Line Filter for Segmentation and Visualization of Curvilinear Structures in Medical Images. In: Troczak, J., Mösges, R., Grimson, W.E.L. (eds.) *CVRMed-MRCAS 1997. LNCS*, vol. 1205, pp. 213–222. Springer, Heidelberg (1997)
25. Schneider, M., Hirsch, S., Weber, B., Székely, G.: Physiologically Based Construction of Optimized 3-D Arterial Tree Models. In: Fichtinger, G., Martel, A., Peters, T. (eds.) *MICCAI 2011, Part I. LNCS*, vol. 6891, pp. 404–411. Springer, Heidelberg (2011)
26. Schneider, M., Sundar, H.: Automatic global vessel segmentation and catheter removal using local geometry information and vector field integration. In: *ISBI 2010*, pp. 45–48 (April 2010)
27. Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* 31(3), 1116–1128 (2006),  
<http://www.itksnap.org>